# 改进YOLOv4的温室环境下草莓生育期识别方法

龙洁花[1,2]，郭文忠[1]，林　森[1*]，文朝武[1]，张　宇[1]，赵春江[1]

（1. 北京市农林科学院智能装备技术研究中心，北京 100097；2. 上海海洋大学 信息学院，上海 201306）

**摘　要：** 针对目前设施农业数字化栽培调控技术中对作物的生育期实时检测与分类问题，提出一种改进YOLOv4的温室环境下草莓生育期识别方法。该方法将注意力机制引入到YOLOv4主干网络的跨阶段局部残差模块（Cross Stage Partial Residual，CSPRes）中，融合草莓不同生长时期的目标特征信息，同时降低复杂背景的干扰，提高模型检测精度的同时保证实时检测效率。以云南地区的智能设施草莓为试验对象，结果表明，本研究提出的YOLOv4-CBAM（YOLOv4-Convolutional Block Attention Module）模型对开花期、果实膨大期、绿果期和成熟期草莓的检测平均精度（Average Precision，AP）分别为92.38%、82.45%、68.01%和92.31%，平均精度均值（Mean Average Precision，mAP）为83.79%，平均交并比（Mean Inetersection over Union，mIoU）为77.88%，检测单张图像时间为26.13 ms。YOLOv4-CBAM模型检测草莓生育期的*mAP*相比YOLOv4、YOLOv4-SE、YOLOv4-SC模型分别提高8.7%、4.82%和1.63%。该方法可对草莓各生育期目标进行精准识别和分类，并为设施草莓栽培的信息化、规模化调控提供有效的理论依据。

**关键词：** 目标检测；草莓；生育期识别；YOLOv4；残差模块；注意力机制；损失函数

**中图分类号：** S126　　**文献标志码：** A　　**文章编号：** 202109-SA006

引用格式：龙洁花，郭文忠，林森，文朝武，张宇，赵春江．改进YOLOv4的温室环境下草莓生育期识别方法[J]. 智慧农业(中英文), 2021, 3(4): 99-110.

LONG Jiehua, GUO Wenzhong, LIN Sen, WEN Chaowu, ZHANG Yu, ZHAO Chunjiang. Strawberry growth period recognition method under greenhouse environment based on improved YOLOv4[J]. Smart Agriculture, 2021, 3(4): 99-110. (in Chinese with English abstract)



## 1　引　言

中国是世界上草莓生产和消费的第一大国，草莓生产是促进中国农民增收致富的重要经济作物[1]。草莓生育期可以作为灌溉、施肥、环境控制等智能管理的决策依据[2]。目前，草莓生育期的识别主要依靠人工观测，工作效率低，无法满足实时、快速的监测需求。近年，深度学习在作物特征识别方面开始应用，由于温室环境复杂，草莓各个生长时期之间存在密集分布、叶片遮挡、果实重叠等因素，给草莓生育期识别带来了困难。因此，研究一种温室环境下草莓生育期识别方法对提高水肥一体化、温室环控系统、机器人等智能装备作业精度具有重要意义。

近年来，随着深度学习技术在场景识别、物体分类等方面的研究越来越成熟[3]，其对果蔬的识别也逐渐成为国内外研究的热点[4,5]。深度学

收稿日期：2021-09-14　修订日期：2021-11-08

基金项目：北京市科技计划（Z211100004621006）；北京市农林科学院青年基金（QNJJ202027）；宁夏回族自治区重点研发计划项目（2018BBF02024）

作者简介：龙洁花（1997—），女，硕士研究生，研究方向为农业机器人和机器视觉。E-mail：seven060422@163.com。

*通信作者：林　森（1984—），男，高级工程师，研究方向为农业智能装备。电话：15652795990。E-mail：linseng@nercita.org.cn。

习中目标检测模型主要分为两类。一类是以R-CNN（Region-Convolutional Neural Network）、Faster R-CNN和Mask R-CNN为代表的基于候选区域的两阶段目标检测方法，该类方法首先产生目标的候选区域，然后利用卷积网络提取区域特征，对候选区域中的目标进行分类和回归，识别精度高但速度较慢，难以满足实时检测要求[6]。Lin和Chen[7]提出了一种基于区域的目标检测方法Faster R-CNN对室外草莓花朵进行检测，检测精度为86.1%，但检测时间达0.118 s。Yu等[8]提出一种改进的Mask R-CNN对成熟和未成熟草莓进行分割，平均检测精度为95.78%，但检测速度仅为8 f/s。可见基于两阶段目标检测方法检测速度有待进一步提高。另一类是以SSD（Single Shot MultiBox Detector）、YOLO（You Only Look Once）系列为代表的基于无侯选区域的单阶段目标检测方法，可直接通过网络产生目标类别和边界框而无需选择候选区域，此类方法检测速度快且识别准确率较高，可满足实时检测要求[9]。刘小刚等[10]提出一种改进的YOLOv3算法对复杂环境下成熟和未成熟草莓静态图像和动态视频进行识别，对成熟和未成熟草莓的识别准确率分别为97.14%和96.51%，且每张图片平均检测时间为35.99 ms。赵春江等[11]提出一种级联卷积神经网络对温室环境下番茄不同花期进行检测，该方法首先采用特征金字塔网络（Feature Pyramid Network，FPN）分割出番茄花束区域，随后将分割后的花束传入YOLOv3网络提取番茄不同花期的小目标特征，以实现花期识别，检测时间为12.54 ms。刘天真等[12]提出一种融入SE（Squeeze-and-Excitation）注意力机制（Attention Mechanism）的YOLOv3网络对复杂环境下冬枣果实进行识别，通过在YOLOv3主干网络的最后两个残差块后嵌入SE模块，增强特征表达能力，试验表明融入SE的YOLOv3模型平均检测精度为82.01%，检测耗时0.0723 s。可见基于单阶段目标检测的YOLO算法在复杂环境下具有较好的识别性能和实时性。

目前研究大多只对作物成熟度或者花朵进行识别，对作物的花期和果实生长期同时识别的研究较少。本研究以不同时期的草莓图像为研究对象，提出一种改进YOLOv4的温室环境下草莓生育期识别方法，将注意力机制引入YOLOv4主干网络的跨阶段局部残差模块中，增加识别草莓不同生长时期目标的特征，同时降低背景信息的干扰，以提高温室环境下草莓生育期检测精度，为草莓智能化生产管控等提供决策依据。

## 2 材料与方法

### 2.1 样本采集及预处理

#### 2.1.1 样本采集

草莓数据采集于云南省昆明市富民县种植基地，品种为章姬，采用手机萤石云软件手动远程采集草莓图片，镜头分辨率为1920×1080 px，如图1所示。样本在2020年8月13日到2020年11月13日期间采集，历经3个月，分别在每天的早上、中午、下午采集一次样本数据，每次采集2张图片，分别采集不同生长阶段、不同遮挡程度的草莓样本图片，经筛选后共300张。为考虑模型训练图片大小对计算机的性能要求，将图片大小统一压缩为960×540 px，提高网络处理效率和小目标检测的实时性[13]。根据图片采集情况将草莓样本分为4个时期：开花期、果实膨大期、绿果期和成熟期，如图2所示。其中开花期84张，果实膨大期58张，绿果期33张，成熟期87张，包含多种时期的图片38张。

#### 2.1.2 样本增强

为提高网络模型泛化能力和鲁棒性，采用数据增强方法增加草莓生育期样本数量[14]，防止网络因训练样本不足导致过拟合。采用左右翻转、调整图像的亮度、对比度，以及增加噪声等方法对草莓生育期样本图片进行数据增强，每张图片增强5次，增强后的草莓样本图片集为1500张，并按照12:2:1的比例将数据集划分为训练集（1200张）、测试集（200张）和验证集（100张）。采用LabelImg标注工具对每个样本进行人



(a)云南草莓种植基地



(b)草莓图像采集设备

图1 草莓种植基地与图像采集设备

Fig. 1 Strawberry planting base and image acquisition equipment



(a)开花期



(b)果实膨大期



(c)绿果期



(d)成熟期

图2 不同生育期草莓样本

Fig. 2 Strawberry sample of different growth periods

工标注，生成一个目标对象二维像素坐标信息的XML文件。

## 2.2 草莓生育期识别模型构建

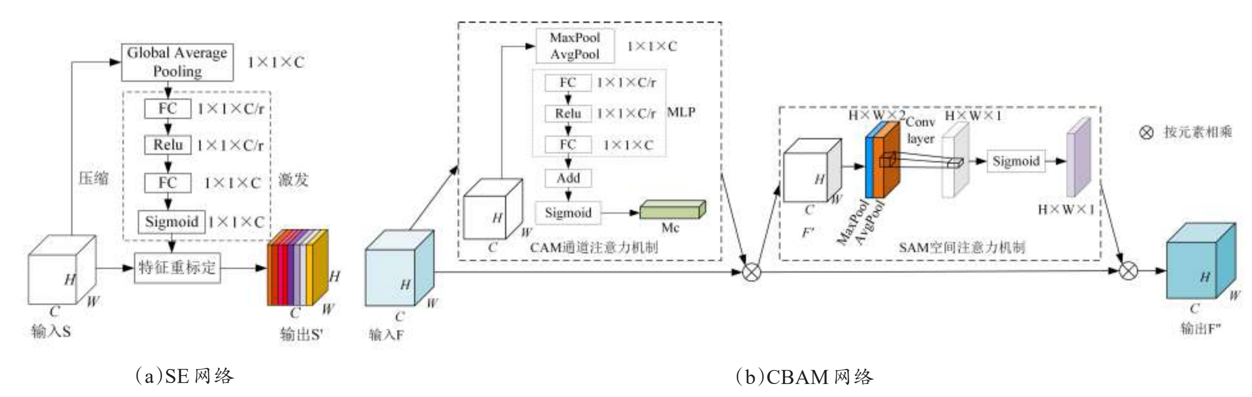
### 2.2.1 注意力机制

近年来，注意力机制在图像处理、语音识别和情感分析等领域具有广泛应用[15]，注意力机制通过对神经网络传播过程中的特征通道加以不同的权重，使得网络更加重视权重较大的通道以进行参数更新[16]，其核心思想在于让神经网络能够忽略无关特征信息而关注重要信息，从而减少任务复杂度，提高检测效率。图像领域的注意力一般集中于提取特征的通道域和实现像素之间的空间域，注意力图与通道维度和空间维度中的特征图相乘提取更加细化的信息特征。

SE（Squeeze-and-Excitation）[17]是一种从通道维度提取特征的卷积神经网络注意力机制，采用特征重标定策略，让网络通过学习的方式自动获取每个特征通道的重要信息[18]，主要包括压缩和激发两个过程，网络结构如图3（a）所示。压缩阶段通过全局平均池化将空间大小为H×W×C的特征图压缩成1×1×C的一维特征向量，随后传入激发阶段。激发阶段主要由2个全连接层（Fully Connected，FC）组成，第1个全连接层有C/r个神经元，输入为1×1×C，输出为1×1×C/r，其中$r$为降维缩放参数，用于压缩全连接层参数；第2个全连接层有C个神经元，输入为1×1×C/r，输出为1×1×C，相比直接使用1个全连接层可更好地拟合通道间复杂的非线性关系，减少模型复杂度。经Sigmoid激活函数得到大小为1×1×C的一维向量，最后经特征重标定步骤，将激发阶段输出的一维向量与原输入特征图S按通道权重相乘得到大小为H×W×C的输出特征图S'。

CBAM（Convolutional Block Attention Module）[19]是一种从通道和空间两个维度提取特征的卷积神经网络注意力机制，网络结构如图3（b）所示，包含CAM（Channel Attention Module）通道注意力机制和SAM（Spartial Attention Module）空间注意力机制。CAM相比SE多了一个并行的最大池化层，首先通过执行平均池化和最大池化将空间大小为H×W×C的特征图压缩为1×1×C的一维向量，得到当前特征图的全局压缩特征量。池化后的一维向量传入多层

感知器区域（Muti-Layer Perception，MLP），MLP主要由2个全连接层构成，第1个全连接层将通道维数从C维降至C/r维，第2个全连接层将通道数从C/r维增加至C维。经全连接层后的特征按元素相加，再进行Sigmoid运算生成大小为1×1×C的一维向量**Mc**。**Mc**与输入大小为H×W×C的特征图F按元素相乘得到大小为H×W×C的特征图F'作为SAM模块的输入。SAM输入特征图F'先分别在通道维度上进行最大池化和平均池化操作得到2个大小为H×W×1的特征图，将这2个特征图按通道维度拼接成H×W×2的特征图，再采用卷积核大小为3×3的卷积层对拼接后的特征图降维至H×W×1的特征图，通过Sigmoid激活函数得到大小为H×W×1的空间注意力特征图，最后将其与SAM模块的输入特征图F'按元素相乘得到大小为H×W×C的输出特征图F"。




(a)SE网络　　(b)CBAM网络

图3　SE和CBAM注意力机制网络结构图

Fig. 3　SE and CBAM attention mechanism network structure diagram

### 2.2.2　改进的YOLOv4网络结构

YOLOv4[20]在YOLOv3[21]基础上采用了近些年神经网络领域中优秀的算法模型和训练技巧[22]，不仅提高了物体检测精度和速度，且对于遮挡的物体检测性能相比YOLOv3更加优越，能满足实时性检测。由于草莓各个生育期为小目标，且所处为非结构化环境，花朵、果实之间存在密集分布、遮挡等因素，网络在前向传播过程中随着网络层数的加深，受遮挡和叶子背景干扰的小目标特征表示愈发减弱，导致这些目标的细节特征在整个深层网络传播过程中消失[23]，造成漏检测或误检测，因此加强小目标特征学习和降低背景干扰极为重要。针对此问题，本研究将注意力机制融入到YOLOv4特征提取网络中的跨阶段局部残差模块（Cross Stage Partial Residual，CSPRes）[24]结构中，通过增加待识别目标区域的特征权值以降低背景信息对识别任务的干扰[25]，从而提高检测精度。改进的YOLOv4网络结构如图4所示，主要由融入注意力机制的特征提取网络、特征融合模块，以及YOLOv3 head这3部分组成。

融入注意力机制的特征提取网络主要由CBM、CSP1-SE/CBAM、CSP2-SE/CBAM、CSP8-SE/CBAM、CSP8-SE/CBAM、CSP4-SE/CBAM组成。CBM主要由卷积（Conv）、批量归一化（Batch Normalization，BN）和Mish激活函数构成，主要用于对特征图进行降维。CSP*X*（Cross Stage Partial *X*）主要由CBM卷积操作和*X*个残差模块（ResUnit）[26]级联构成，卷积操作对特征进行降维，残差模块在浅层网络和深层网络间以跳跃连接的方式将输入直接与输出相加，用于解决深度神经网络中的梯度爆炸问题。CSP*X*-SE/CBAM表示分别将SE通道注意力机制、CBAM空间注意力机制融入到CSPRes结构中用于加强草莓生育期目标特征学习。特征融合模块主要由空间金字塔池化（Spatial Pyramid Pooling，SPP）[27]和路径聚合网络（Path Aggregation Network，PANet）[28]组成。SPP使用不同尺度的

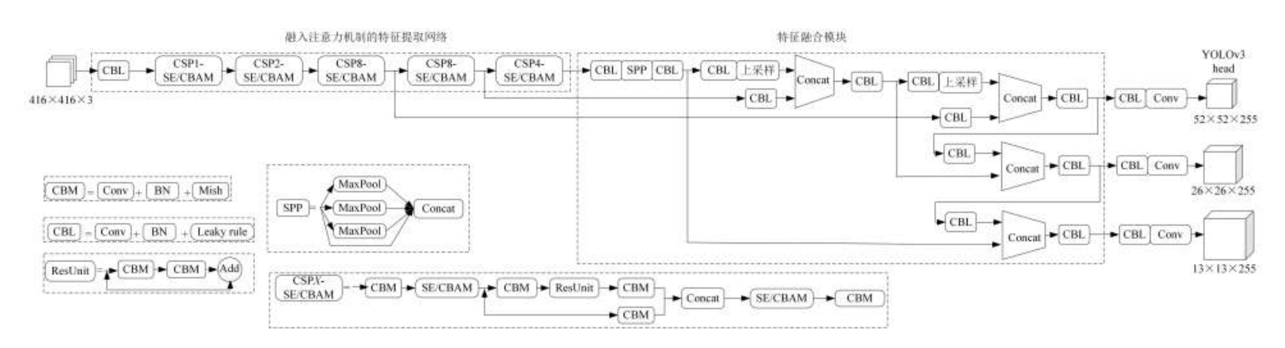



图 4　改进的 YOLOv4 网络结构图

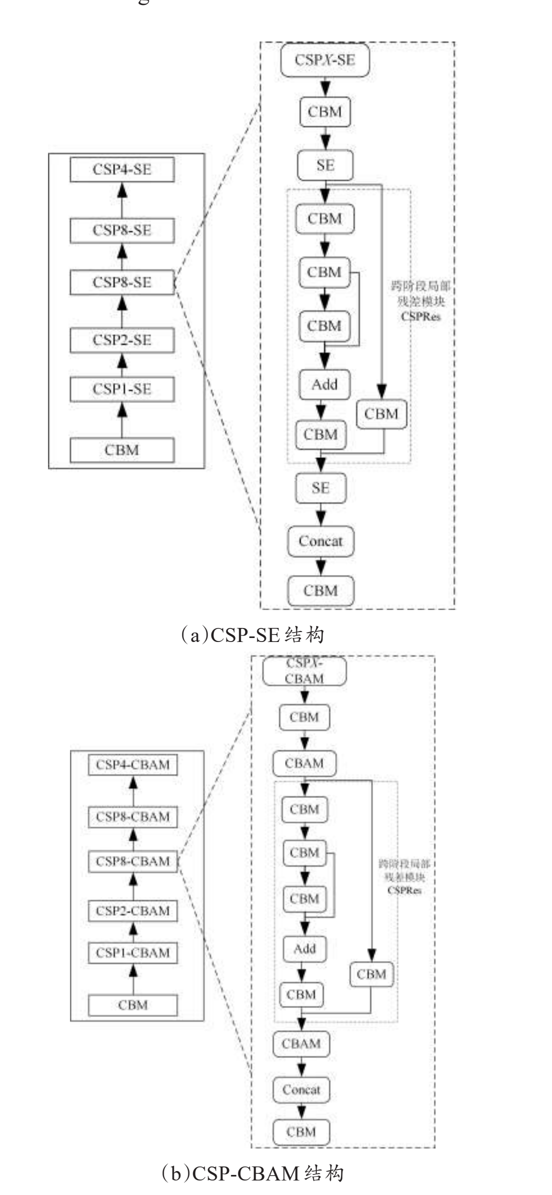
Fig. 4  Improved YOLOv4 network structure diagram

最大池化核对特征图像进行池化操作，用于增加主干网络提取特征的接收范围；PANet将不同检测层参数特征进行聚合，进一步提高特征表达能力。输入图像的大小首先经网络调整为416×416 px，通过融入注意力机制的特征提取网络和特征融合模块得到大小为52×52 px、26×26 px、13×13 px 的目标特征图，YOLOv3 head 对这 3 种不同尺度大小的特征图进行分类和回归预测，输出类别和边界框位置。

### 2.2.3　融入注意力机制的特征提取网络

由于注意力机制模块可以插入到网络任意位置，考虑到不更改 CSPDarknet53 的网络结构而加载预训练权重，本研究分别将 SE 通道注意力机制、CBAM 空间注意力机制嵌/插入到主干网络 CSPDarknet53 的 CSPRes 结构的第一层卷积前和最后一层卷积后，不会改变网络结构。融入 SE 的主干网络将其称为 CSP-SE（Cross Stage Partial-Squeeze and Excitation），融入 CBAM 的主干网络将其称为 CSP-CBAM（Cross Stage Partial-Convolutional Block Attention Module），网络结构分别如图 5（a）和 5（b）所示。图 5 中 CSP*X*-SE、CSP*X*-CBAM 中的 *X* 都表示跨阶段局部残差模块个数，分别为 1、2、8、8 和 4。同时融合注意力机制的主干网络的降维缩放参数 $r$ 参照 SENet 网络，设置 $r$ =16 对全连接层参数进行压缩，用于平衡速度和检测性能。

如图 5 所示，CSP-SE 网络结构由 CBM、CSP1-SE、CSP2-SE、2 个 CSP8-SE 和 CSP4-SE 组成，CSP-CBAM 网络结构由 CBM、CSP1-




（a）CSP-SE 结构

（b）CSP-CBAM 结构

图 5　融入注意力机制的特征提取网络

Fig. 5  Feature extraction network integrated with attention mechanism

CBAM、CSP2-CBAM、2 个 CSP8-CBAM 和 CSP4-CBAM 组成。CSP*X*-SE 和 CSP*X*-CBAM 结构分别将 SE 和 CBAM 注意力机制插入到 CSPRes 的第一层卷积前和最后一层卷积后，网络在训练过程中对经过使用注意力机制的通道维度和空间维度加以更高的特征权重，使得网络更加重视权值较大的特征以进行学习，忽视无关特征信息，以提取更多目标特征传递给后续特征融合模块。

#### 2.2.4 草莓生育期检测模型架构

本研究将融入注意力机制的特征提取网络作为改进 YOLOv4 的主干网络，对草莓各个生育期进行识别，草莓生育期检测模型架构如图 6 所示。主要分为 2 部分：第一部分为数据预处理，第二部分为草莓生育期检测网络。数据预处理在数据采集基础上采用数据增强方法对样本进行扩增，用于提高模型泛化能力。草莓生育期检测网络主要由融入注意力机制的特征提取网络和特征融合模块组成。融入注意力机制的特征提取网络在训练时通过学习忽略无关目标特征的信息而关注重点信息，提取不同时期更准确的草莓特征传入特征融合网络中，特征融合模块对不同尺度的特征进行聚合，得到预测框和草莓类别，分别为开花期、果实膨胀期、绿果期和成熟期。网络在训练过程中学习各个阶段草莓的最重要特征并进行区分，同时根据边界框、类别和置信度损失函数不断地调整模型参数使网络最终达到收敛状态，得到更加准确的类别预测和边界框位置，实现草莓各个生育期的精准检测。

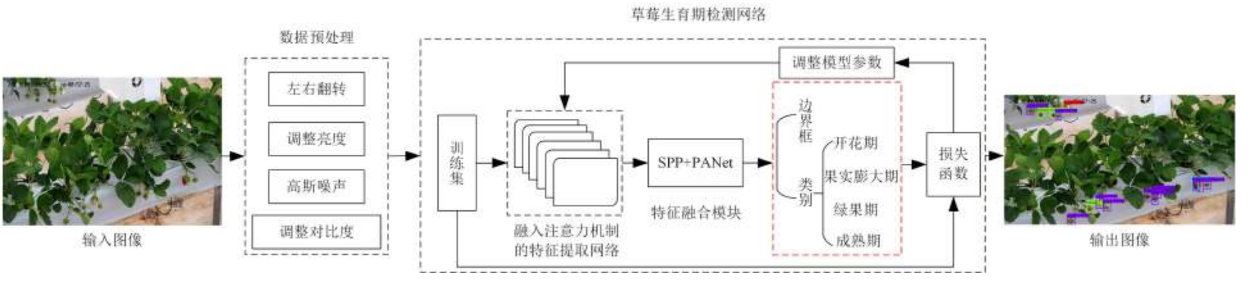



图 6 草莓生育期检测模型架构

Fig. 6 Detection model architecture of strawberry growth period

#### 2.2.5 损失函数

YOLO 损失包含类别损失、边界框位置损失和置信度损失。类别损失和置信度损失都采用二元交叉熵损失函数，而目标检测中的一项重要任务就是确定目标边界框位置。改进的 YOLOv4 网络采用 CIoU Loss（Complete IoU Loss）计算边界框位置损失，同时考虑到了边界框重合度、中心距离和宽高比的信息，进一步提高模型精度。CIoU Loss 计算如公式（1）所示。

$$L_{\mathrm{CIoU}} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \tag{1}$$

其中，$IoU$ 为预测框与目标框的交并比；$b$ 为先验框的中心点；$b^{gt}$ 为目标框的中心点；$\rho$ 为两个中心点的欧式距离；$c$ 为先验框和目标框之间的最小矩形的对角线距离；$\alpha$ 为权衡参数；$v$ 为长宽比一致性衡量参数。$\alpha$ 和 $v$ 计算方法如公式（2）和公式（3）所示。

$$\alpha = \frac{v}{(1 - \mathrm{IoU}) + v} \tag{2}$$

$$v = \frac{4}{\Pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{3}$$

其中，$w$ 为预测框的宽；$h$ 为预测框的高。

## 3 试验与结果分析

### 3.1 试验环境及参数设置

草莓检测网络均在 Pytorch 深度学习框架中训练。硬件环境为 Intel® CoreTM i7-9800X CPU @ 3.8 GHz×16 中央处理器，16 GB 运行内存，12 GB 的 GeForce GTX 1080ti 显卡。软件环境为 Ubuntu16.04 系统，网络在 Anaconda3 虚拟环境下运行，配置安装 python3.7、Cuda10.0 和 Cudnn7.4。图片输入大小为 416×416 px。训练

采用冻结训练策略。首先冻结主干网络参数训练1000步，学习率设置为0.01，随后解冻网络训练2000步，学习率设置为0.001，总迭代步数为3000，采用冻结训练可以加快网络训练速度，也可防止训练初期权重被破坏。IoU阈值设置为0.5。

### 3.2 试验评价指标

本研究采用准确率（Precision，P）、召回率（Recall，R）、平均精度（Average Precision，AP）、平均精度均值（Mean Average Precision，mAP）、平均交并比（Mean Intersection over Union，mIoU）作为网络性能的主要评价指标。其中*mAP*是衡量多类别目标检测模型性能的重要指标，交并比是预测框与真实框的重合程度，用来表示目标定位精度。各评价指标计算如公式（4）～公式（7）所示：

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

$$AP = \int_0^1 \mathrm{P}(R)dR \tag{6}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP \tag{7}$$

其中，*TP*表示模型预测为正的正样本个数，个；*FP*表示模型预测为正的负样本个数，个；*FN*表示模型预测为负的正样本个数，个；*TN*表示模型预测为负的负样本个数，个；*AP*为准确率在召回率上的积分，只计算单一类别精度值；*mAP*为每一个类别*AP*的平均值，其中*i*为类别编号，*n*为类别个数，个。本研究中草莓生育期类别分别为开花期、果实膨大期、绿果期和成熟期，故$n$=4。

使用模型增量参数和平均检测时间作为检测速度的评估指标。

### 3.3 草莓生育期检测性能结果分析

为有效对比YOLOv4网络中融入SE和CBAM注意力机制的性能，将SAM与SE注意力机制级联组合成先SAM空间后SE通道的混合注意力机制，称其为SC（Spatial and Channel）注意力机制模块。同时将SC注意力机制融入到YOLOv4主干网络的跨阶段局部残差模块的第一层卷积前和最后一层卷积后，融入SC的主干网络称其为CSP-SC（Cross Stage Partial-Spatial and Channel）。分别将CSP-SE、CSP-CBAM、CSP-SC作为改进YOLOv4的主干网络，记为YOLOv4-SE、YOLOv4-CBAM、YOLOv4-SC。为验证融入注意力机制的改进YOLOv4模型在温室环境下检测草莓生育期的性能，与YOLOv4网络在测试集上对草莓开花期、果实膨大期、绿熟期和成熟期这4个生长时期检测的性能进行对比分析，结果如表1所示。

**表1 不同检测模型在草莓生育期测试集上的性能测试结果**

Table 1 Performance testing results of different detection models on the strawberry growth period test set

| 模型 | 参数增量 | 不同生长时期草莓检测精度*AP*/% | | | | *mAP*/% | *mIoU*/% | 平均检测时间/ms |
|---|---|---|---|---|---|---|---|---|
| | | 开花期 | 果实膨大期 | 绿果期 | 成熟期 | | | |
| YOLOv4 | 0 | 77.96 | 77.10 | 54.77 | 90.53 | 75.09 | 72.35 | 25.00 |
| YOLOv4-SE | 856,576 | 86.29 | 79.17 | 59.41 | 91.01 | 78.97 | 74.42 | 25.45 |
| YOLOv4-SC | 856,990 | 90.34 | 81.06 | 65.25 | 91.98 | 82.16 | 75.15 | 25.87 |
| YOLOv4-CBAM | 856,990 | 92.38 | 82.45 | 68.01 | 92.31 | 83.79 | 77.88 | 26.13 |

由表1可知，YOLOv4-CBAM对不同生长时期草莓检测的*mAP*和*mIoU*相比YOLOv4分别提高8.7%和5.53%，平均检测时间相比YOLOv4增加1.13 ms；YOLOv4-SC对不同生长时期草莓检测的*mAP*和*mIoU*相比YOLOv4分别提高7.07%和2.8%，平均检测相比YOLOv4增加0.87 ms；YOLOv4-SE对不同生长时期草莓检测的*mAP*和*mIoU*相比YOLOv4分别提高3.88%和2.07%，平均检测时间相比YOLOv4增加0.45 ms。可知在YOLOv4主干网络中加入注意力机制虽然增加了

模型参数数量，但模型检测时间只是略有增加，而YOLOv4-CBAM、YOLOv4-SC和YOLOv4-SE检测性能相比YOLOv4却有显著提升。其次，YOLOv4-CBAM的*mAP*和*mIoU*相比YOLOv4-SE模型分别提高4.82%和3.46%，平均检测时间相比YOLOv4-SE增加0.68 ms，是因为CBAM在网络结构上比SE多了一个空间注意力模块，是从通道和空间两个维度提取草莓各个生长时期的特征，使得YOLOv4-CBAM模型相比YOLOv4-SE更加关注于识别物体，从而提高了检测精度和定位精度。YOLOv4-CBAM的*mAP*和*mIoU*相比YOLOv4-SC模型分别提高1.63%和2.73%，平均检测时间相比YOLOv4-SC增加0.26 ms，是因为SC模块中的SE结构相比CBAM中的CAM结构少了一个并行的最大池化层，采用并行的最大池化和平均池化组合相比，使用一个平均池化丢失的特征信息更少，得到的目标特征更丰富，定位准确度更高。综合权衡检测精度和网络运行时间，YOLOv4-CBAM模型检测不同生长时期草莓更具鲁棒性，满足实时性检测要求。

### 3.4 草莓生育期识别效果对比分析

本研究分别采用YOLOv4-CBAM、YOLOv4-SE、YOLOv4-SC和YOLOv4基于草莓图片（随机选取样例1和样例2）对草莓生育期行预测，结果如图7所示。从图7中可以看到，融入注意力机制的改进YOLOv4模型对复杂环境下不同时期的草莓均具有很好的识别效果，且能准确定位不同时期的草莓目标。

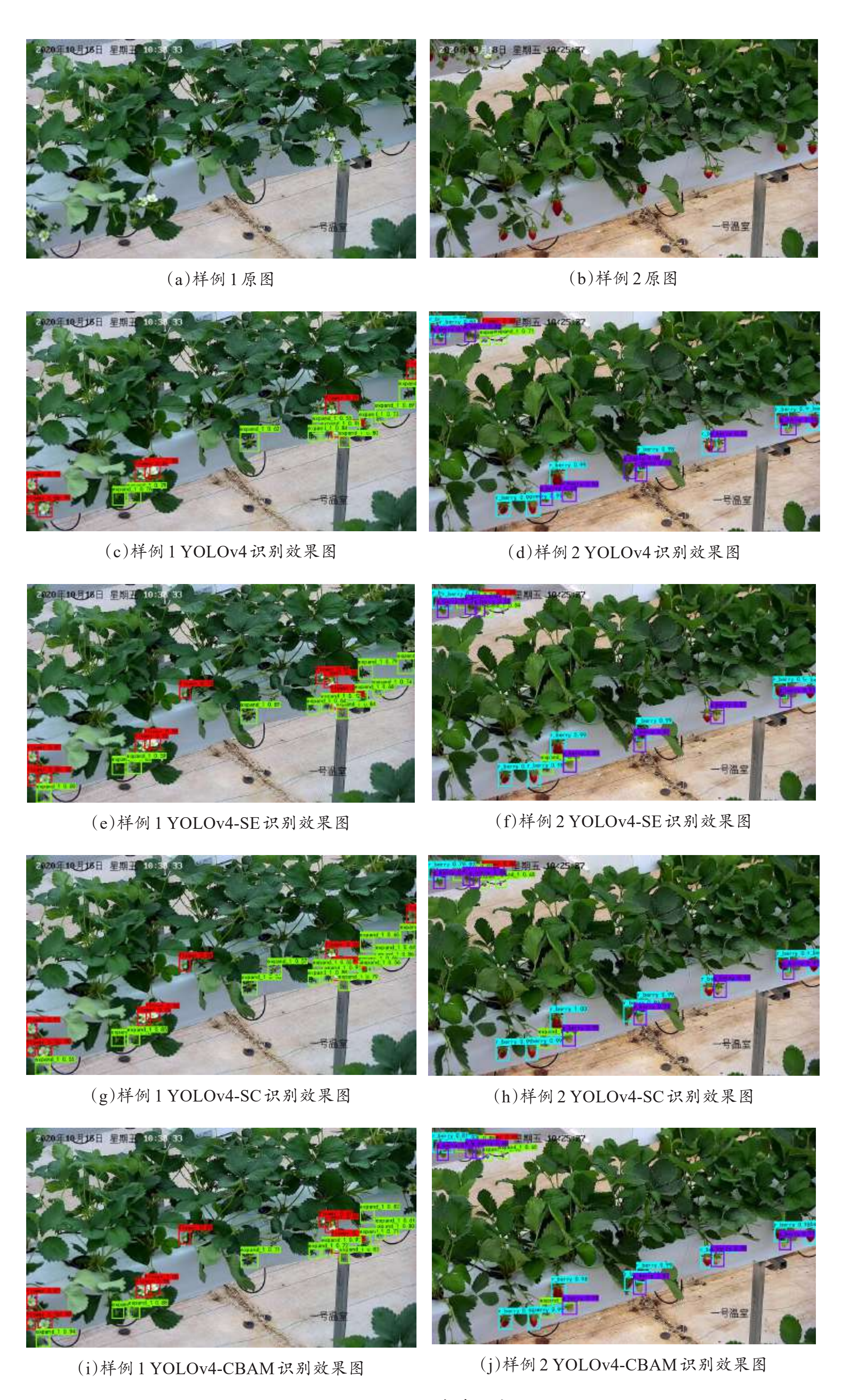
为有效对比不同模型的识别效果，对草莓各个时期进行计数，并统计其置信度大小，结果如表2所示，包括不同模型识别样例1和样例2中不同生长时期草莓的置信度和个数结果。

从表2（108页）可知，YOLOv4-CBAM对开花期、果实膨大期、绿果期和成熟期草莓识别的置信度均高于其他模型，表明YOLOv4-CBAM预测目标正确概率更大。样例1中YOLOv4-CBAM、YOLOv4-SC和YOLOv4-SE识别开花期和果实膨大期数量明显多于YOLOv4，说明增加注意力机制能增加小目标特征提取的能力。样例2中YOLOv4识别绿果期时存在误识别现象，将果实膨大期和成熟期误识别为绿果期，绿果期和果实膨大期之间特征较为相似，被叶子遮挡的成熟期特征不明显，导致YOLOv4出现误识别为绿果期现象，且YOLOv4识别果实膨大期数量少于YOLOv4-SE、YOLOv4-SC和YOLOv4-CBAM模型，是因为果实膨大期颜色特征与叶子背景特征相似，导致YOLOv4识别果实膨大期特征时易受叶子背景干扰，而加入注意力机制可以降低背景信息和其他特征信息的干扰。YOLOv4-SE、YOLOv4-SC和YOLOv4-CBAM模型均无误识别现象，但YOLOv4-CBAM模型相比YOLOv4-SE和YOLOv4-SC模型识别不同生长时期草莓的置信度更高，且YOLOv4-CBAM识别果实膨大期和成熟期个数高于YOLOv4-SE，说明YOLOv4-CBAM模型能提取更加全面丰富的特征，更加关注于识别目标，从而提高检测准确率。

## 4 结 论

本研究提出了一种融合注意力机制的改进YOLOv4模型对不同生育期的草莓进行识别，将注意力机制融入到YOLOv4主干网络的CSPRes的第一层卷积前和最后一层卷积后，网络通过注意力机制对重要的目标特征加以更高的特征权重，以融合草莓不同生育期的特征信息同时降低复杂背景信息的干扰，提高检测精度。

（1）在草莓生育期测试集上试验结果表明，提出的YOLOv4-CBAM模型对草莓开花期、果实膨大期、绿果期和成熟期的检测平均精度分别为92.38%、82.45%、68.01%和92.31%，平均精度均值为83.79%，平均交并比为77.88%，平均检测时间为26.13 ms。与其他融入注意力机制的模型相比，YOLOv4-CBAM模型综合权衡了检测精度与网络运行速度，且定位精度最高，可满足实时检测草莓生育期状态需求。



(a)样例1原图　(b)样例2原图

(c)样例1 YOLOv4识别效果图　(d)样例2 YOLOv4识别效果图

(e)样例1 YOLOv4-SE识别效果图　(f)样例2 YOLOv4-SE识别效果图

(g)样例1 YOLOv4-SC识别效果图　(h)样例2 YOLOv4-SC识别效果图

(i)样例1 YOLOv4-CBAM识别效果图　(j)样例2 YOLOv4-CBAM识别效果图

图7　不同模型识别草莓生育期效果图

Fig. 7 Effect diagram of strawberry growth period identification by different models

**表2 不同模型识别不同生长时期草莓的置信度和个数结果**

Table 2 Different models recognize the confidence and number of strawberries in different growth periods

| 模型 | 样例1 | | | 样例2 | | |
|---|---|---|---|---|---|---|
| | 生长期 | 置信度 | 数量/个 | 生长期 | 置信度 | 数量/个 |
| YOLOv4 | 开花期 | 0.98、0.98、0.92、0.91、0.86、0.81、0.75、0.64 | 8 | 开花期 | 0.88 | 1 |
| | | | | 果实膨大期 | 0.87、0.71 | 2 |
| | 果实膨大期 | 0.96、0.84、0.84、0.80、0.79、0.76、0.73、0.69、0.63、0.55 | 10 | 绿果期 | 0.93、0.91、0.85、0.72、0.69、0.59、0.55、0.52 | 8* |
| | | | | 成熟期 | 0.99、0.99、0.99、0.98、0.95、0.94、0.90、0.85 | 9 |
| YOLOv4-SE | 开花期 | 1.00、0.99、0.95、0.92、0.90、0.87、0.86、0.73、0.53、0.50 | 10 | 开花期 | 0.89 | 1 |
| | | | | 果实膨大期 | 0.84、0.79、0.54 | 3 |
| | 果实膨大期 | 0.92、0.89、0.84、0.83、0.80、0.80、0.79、0.74、0.68、0.64、0.59 | 11 | 绿果期 | 0.96、0.94、0.87、0.87、0.84、0.81、0.56 | 7 |
| | | | | 成熟期 | 0.99、0.99、0.99、0.98、0.98、0.98、0.91、0.85 | 8 |
| YOLOv4-SC | 开花期 | 1.00、0.98、0.96、0.95、0.91、0.91、0.79、0.76、0.58、0.54 | 10 | 开花期 | 0.88 | 1 |
| | | | | 果实膨大期 | 0.97、0.92、0.89、0.65 | 4 |
| | 果实膨大期 | 0.96、0.96、0.96、0.88、0.85、0.84、0.79、0.66、0.65、0.64、0.57 | 11 | 绿果期 | 0.97、0.97、0.95、0.95、0.91、0.90、0.86 | 7 |
| | | | | 成熟期 | 1.00、0.99、0.99、0.99、0.97、0.96、0.92、0.81、0.76、0.72 | 10 |
| YOLOv4-CBAM | 开花期 | 1.00、1.00、0.96、0.96、0.95、0.87、0.81、0.79、0.76、0.61 | 10 | 开花期 | 0.93 | 1 |
| | | | | 果实膨大期 | 0.98、0.95、0.81、0.65 | 4 |
| | 果实膨大期 | 0.97、0.95、0.94、0.85、0.83、0.82、0.80、0.72、0.71、0.71、0.61 | 11 | 绿果期 | 0.99、0.98、0.98、0.97、0.94、0.88、0.87 | 7 |
| | | | | 成熟期 | 0.99、0.99、0.99、0.98、0.98、0.98、0.93、0.81、0.80、0.74 | 10 |

注：*表示YOLOv4出现误识别现象，将果实膨大期和成熟期误识别为绿果期

（2）融入注意力机制的改进YOLOv4模型对不同生育期草莓的检测精度、平均交并比均高于YOLOv4模型，可知融入注意力机制有助于网络学习更加重要的目标特征，忽略无关特征，从而降低背景信息干扰。YOLOv4-CBAM模型的平均检测时间相比YOLOv4、YOLOv4-SE和YOLOv4-SC模型略有增加，但对实时性检测影响不大。针对此问题，下一步研究可构建更加轻量级的CBAM模型嵌入到YOLOv4网络中，有效提高模型运行速度的同时提高检测精度。

## 参考文献：


[1] 张更, 颜志明, 王全智, 等. 我国设施草莓无土栽培技术的研究进展与发展建议[J]. 江苏农业科学, 2019, 47(18): 58-61.
ZHANG G, YAN Z, WANG Q, et al. Research progress and development suggestions of soilless culture techniques of China's facility strawberry[J]. Jiangsu Agricultural Sciences, 2019, 47(18): 58-61.

[2] 林森, 郭文忠, 郑建锋, 等. 基于知识图谱和机器视觉的智慧草莓生产托管服务系统实践[J]. 农业工程技术, 2021, 41(4): 17-20.
LIN S, GUO W, ZHENG J, et al. Practice of smart strawberry production hosting service system based on knowledge graph and machine vision[J]. Agricultural Engineering Technology, 2021, 41(4): 17-20.

[3] 徐建鹏, 王杰, 徐祥, 等. 基于RAdam卷积神经网络的水稻生育期图像识别[J]. 农业工程学报, 2021, 37(8): 143-150.
XU J, WANG J, XU X, et al. Image recognition for different developmental stages of rice by RAdam deep convolutional neural networks[J]. Transactions of the CSAE, 2021, 37(8): 143-150.

[4] CHEN W, LU S, LIU B, et al. Detecting citrus in or-

chard environment by using improved YOLOv4[J]. Scientific Programming, 2020(1): 1-13.

[ 5 ] PÉREZ-BORRERO I, MARÍN-SANTOS D, GEGÚNDE-ARIAS M E, et al. A fast and accurate deep learning method for strawberry instance segmentation[J]. Computers and Electronics in Agriculture, 2020, 178(6): 105736-105748.

[ 6 ] 刘芳, 刘玉坤, 林森, 等. 基于改进型YOLO的复杂环境下番茄果实快速识别方法[J]. 农业机械学报, 2020, 51(6): 229-237.
LIU F, LIU Y, LIN S, et al. Fast recognition method for tomatoes under complex environments based on improved YOLO[J]. Transactions of the CSAM, 2020, 51(6): 229-237.

[ 7 ] LIN P, CHEN Y. Detection of strawberry flowers in outdoor field by deep neural network[C]// 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). Piscataway, New York, USA: IEEE, 2018: 482-486.

[ 8 ] YU Y, ZHANG K, YANG L, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN[J]. Computers and Electronics in Agriculture, 2019(163): 104846-104855.

[ 9 ] 李志军, 杨圣慧, 史德帅, 等. 基于轻量化改进YOLOv5的苹果树产量测定方法[J]. 智慧农业, 2021, 3(2): 100-114.
LI Z, YANG S, SHI D, et al. Yield estimation method of apple tree based on improved lightweight YOLOv5[J]. Smart Agriculture, 2021, 3(2): 100-114.

[10] 刘小刚, 范诚, 李加念, 等. 基于卷积神经网络的草莓识别方法[J]. 农业机械学报, 2020, 51(2): 237-244.
LIU X, FANG C, LI J, et al. Identification method of strawberry based on convolutional neural network[J]. Transactions of the CSAM, 2020, 51(2): 237-244.

[11] 赵春江, 文朝武, 林森, 等. 基于级联卷积神经网络的番茄花期识别检测方法[J]. 农业工程学报, 2020, 36(24): 143-152.
ZHAO C, WEN C, LIN S, et al. Tomato florescence recognition and detection method based on cascaded neural network[J]. Transactions of the CSAE, 2020, 36(24): 143-152.

[12] 刘天真, 滕桂法, 苑迎春, 等. 基于改进YOLOv3的自然场景下冬枣果实识别方法[J]. 农业机械学报, 2021, 52(5): 17-25.
LIU T, TENG G, YUAN Y, et al. Winter jujube fruit recognition method based on improved YOLOv3 under natural scene[J]. Transactions of the CSAM, 2021, 52(5): 17-25.

[13] 刘立波, 程晓龙, 赖军臣. 基于改进全卷积网络的棉田冠层图像分割方法[J]. 农业工程学报, 2018, 34(12): 193-201.
LIU L, CHENG X, LAI J, et al. Segmentation method for cotton canopy image based on improved fully convolutional network model[J]. Transactions of the CSAE, 2018, 34(12): 193-201.

[14] 朱逢乐, 郑增威. 基于图像和卷积神经网络的蝴蝶兰种苗生长势评估[J]. 农业工程学报, 2020, 36(9): 185-194.
ZHU F, ZHENG Z. Image-based assessment of growth vigor for *Phalaenopsis aphrodite* seedlings using convolutional neural network[J]. Transactions of the CSAE, 2020, 36(9): 185-194.

[15] CHAUDHARI S, MITHAL V, POLATKAN G, et al. An attentive survey of attention models[J/OL]. arXiv: 1904.02874v3[cs.Lg]. 2021.

[16] 徐诚极, 王晓峰, 杨亚东. Attention-YOLO: 引入注意力机制的YOLO检测算法[J]. 计算机工程与应用, 2019, 55(6): 13-23.
XU C, WANG X, YANG Y. Attention-YOLO: YOLO detection algorithm that introduces attention mechanism[J]. Computer Engineering and Applications, 2019, 55(6): 13-23.

[17] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[C]// The IEEE conference on computer vision and pattern recognition. Piscataway, New York, USA: IEEE, 2018: 2011-2023.

[18] 汶茂宁. 基于轮廓波CNN和选择性注意机制的高分辨SAR目标检测和分类[D]. 西安: 西安电子科技大学, 2018.
WEN M. Target detection and classification for high-resolution SAR image based on contourlet CNN and selective attention mechanism[D]. Xi'an: Xidian University, 2018.

[19] WOO S, PARK J, LEE J-Y, et al. CBAM: Convolutional block attention module[C]// The European Conference on Computer Vision(ECCV). Berlin, German: Springer, 2018: 3-19.

[20] BOCHKOVSKIY A, WANG C, LIAO H M. YOLOv4: Optimal speed and accuracy of object detection[J/OL]. arXiv:2004.10934 [cs.CV]. 2020.

[21] REDMON J, FARHAD A. YOLOv3: An incremental improvement[J/OL]. arXiv:1804.02767 [cs.CV]. 2018.

[22] 蒋镕圻, 彭月平, 谢文宣, 等. 嵌入scSE模块的改进YOLOv4小目标检测算法[J]. 图学学报, 2021, 42(4): 546-555.
JIANG R, PENG Y, XIE W, et al. Improved YOLOv4 small target detection algorithm with embedded scSE module[J]. Journal of Graphics, 2021, 42(4): 546-555.

[23] 李文涛, 张岩, 莫锦秋, 等. 基于改进YOLOv3-tiny的田间行人与农机障碍物检测[J]. 农业机械学报,

2020, 51(S1):8-15, 40.
LI W, ZHANG Y, MO J, et al. Detection of pedestrian and agricultural vehicles in field based on improved YOLOv3 tiny[J]. Transactions of the CSAM, 2020, 51 (S1): 8-15, 40.

[24] WANG C, LIAO H M, WU Y, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, New York, USA: IEEE, 2020: 1571-1580.

[25] 温长吉, 娄月, 张笑然, 等. 基于改进稠密胶囊网络模型的植物识别方法[J]. 农业工程学报, 2020, 36(8): 143-155.
WEN C, LOU Y, ZHANG X, et al. Plant recognition method based on an improved dense CapsNet[J]. Transactions of the CSAE, 2020, 36(8): 143-155.

[26] HE K, ZHANG X, REN S, et al. Deep residual Learning for image recognition[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2016: 770-778.

[27] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(9): 1904-1916.

[28] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New York, USA: IEEE, 2018: 8759-8768.



# Strawberry Growth Period Recognition Method Under Greenhouse Environment Based on Improved YOLOv4

LONG Jiehua[1,2], GUO Wenzhong[1], LIN Sen[1*], WEN Chaowu[1], ZHANG Yu[1], ZHAO Chunjiang[1]

（1. *Beijing Academy of Agriculture and Forestry Sciences Intelligent Equipment Technology Research Center*, *Beijing* 100097, *China*; 2. *College of Information Science*, *Shanghai Ocean University*, *Shanghai* 201306, *China*）

**Abstract:** Aiming at the real-time detection and classification of the growth period of crops in the current digital cultivation and regulation technology of facility agriculture, an improved YOLOv4 method for identifying the growth period of strawberries in a greenhouse environment was proposed. The attention mechanism into the Cross Stage Partial Residual (CSPRes) module of the YOLOv4 backbone network was introduced, and the target feature information of different growth periods of strawberries while reducing the interference of complex backgrounds was integrated, the detection accuracy while ensured real-time detection efficiency was improved. Took the smart facility strawberry in Yunnan province as the test object, the results showed that the detection accuracy (AP) of the YOLOv4-CBAM model during flowering, fruit expansion, green and mature period were 92.38%, 82.45%, 68.01% and 92.31%, respectively, the mean average precision (mAP) was 83.78%, the mean inetersection over union (mIoU) was 77.88%, and the detection time for a single image was 26.13 ms. Compared with the YOLOv4-SC model, *mAP* and *mIoU* were increased by 1.62% and 2.73%, respectively. Compared with the YOLOv4-SE model, *mAP* and *mIOU* increased by 4.81% and 3.46%, respectively. Compared with the YOLOv4 model, *mAP* and *mIOU* increased by 8.69% and 5.53%, respectively. As the attention mechanism was added to the improved YOLOv4 model, the amount of parameters increased, but the detection time of improved YOLOv4 models only slightly increased. At the same time, the number of fruit expansion period recognized by YOLOv4 was less than that of YOLOv4-CBAM, YOLOv4-SC and YOLOv4-SE, because the color characteristics of fruit expansion period were similar to those of leaf background, which made YOLOv4 recognition susceptible to leaf background interference, and added attention mechanism could reduce background information interference. YOLOv4-CBAM had higher confidence and number of identifications in identifying strawberry growth stages than YOLOv4-SC, YOLOv4-SE and YOLOv4 models, indicated that YOLOv4-CBAM model can extract more comprehensive and rich features and focus more on identifying targets, thereby improved detection accuracy. YOLOv4-CBAM model can meet the demand for real-time detection of strawberry growth period status.

**Key words:** object detection; strawberry; growth period recognition; YOLOv4; residual module; attention mechanism; loss function